



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

**TESIS**

# **IDENTIFIKASI MAHASISWA YANG MEMPUNYAI KECENDERONGAN LULUS TIDAK TEPAT WAKTU PADA PROGRAM STUDI MMT-ITS DENGAN MENGGUNAKAN ALGORITMA C4.5**

**Oleh : Amelia Halim (9112205304)**

**Dosen Pembimbing : Prof. Dr. Ir. Joko Lianto Buliali, M.Sc.**



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# PENDAHULUAN





# Latar belakang

- Mahasiswa salah satu elemen penting pada Perguruan Tinggi.
- Kualitas Perguruan Tinggi diukur dari mahasiswa dan lulusan.
- Cara meningkatkan level mutu dari Perguruan Tinggi
- Data Mining sebagai pengolah informasi



# Rumusan Masalah

- Faktor-faktor apa yang mempengaruhi kecenderungan mahasiswa lulus tidak tepat waktu?
- Bagaimana membuat suatu aplikasi yang yang mampu mendeteksi kecenderungan calon mahasiswa lulus tidak tepat waktu?





**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# Tujuan Penelitian

- Menganalisa faktor – faktor yang mempengaruhi kecenderungan mahasiswa lulus tidak tepat waktu.
- Membuat aplikasi yang mampu mendeteksi kecenderungan calon mahasiswa lulus tidak tepat waktu.



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# Batasan Masalah

- Memberikan peringatan dini tentang kecenderungan mahasiswa lulus tidak tepat waktu
- Tidak menanggapi pemberian solusi manajerial untuk mengatasi permasalahan mahasiswa



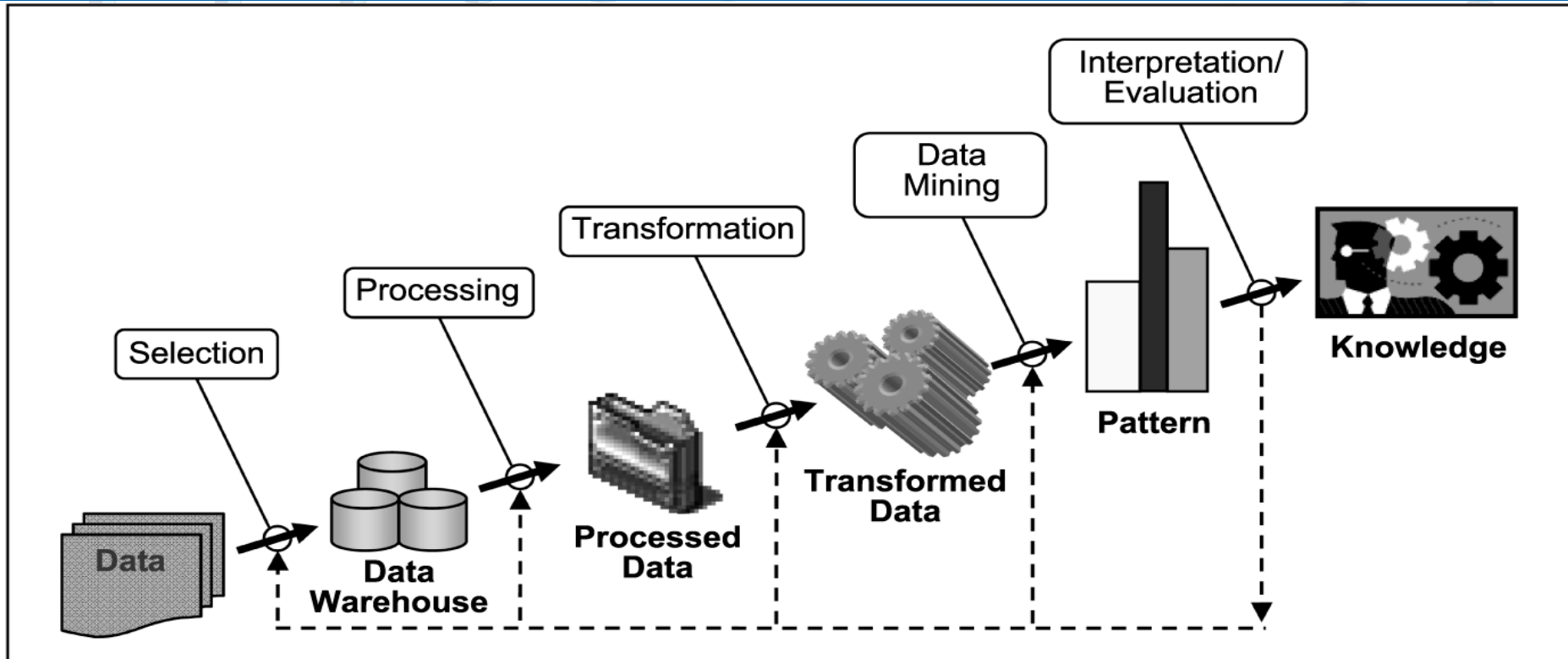


**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# LANDASAN TEORI



# Knowledge Discovery in Databases (KDD)



Knowledge discovery (data mining) in databases (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data





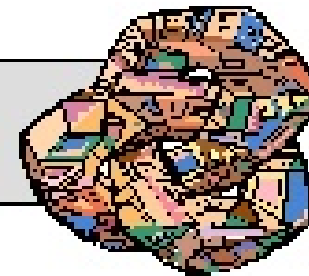
**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# Preprocessing Data



Data Cleaning

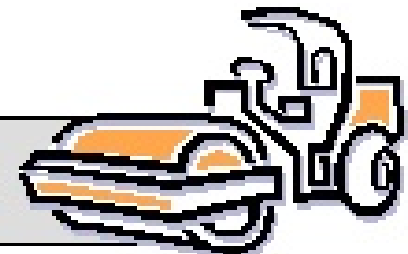
Data Integration



Data Transformation



Data Reduction





# Konversi Data Numerik

IPK S1	
Range	Kelas
4.00 - 3.79	A
3.78 - 3.58	B
3.57 - 3.37	C
3.36 - 3.16	D
< 3.16	E

MBID	
Range	Kelas
100 - 86	A
85 - 72	B
71 - 58	C
57 - 44	D
< 44	E

IPS S1	
Range	Kelas
4.00 - 3.68	A
3.67 - 3.36	B
3.35 - 3.04	C
3.03 - 2.72	D
< 2.72	E

SAKhir	
Range	Kelas
100 - 92	A
91 - 84	B
83 - 76	C
75 - 68	D
< 68	E

Wawancara	
Range	Kelas
100 - 92	A
91 - 84	B
83 - 76	C
75 - 68	D
< 68	E

IPS S2	
Range	Kelas
4.00 - 3.6	A
3.5 - 3.2	B
3.1 - 2.8	C
2.7 - 2.4	D
< 2.4	E

GMAT	
Range	Kelas
100 - 86	A
85 - 72	B
71 - 58	C
57 - 44	D
< 44	E

TOEFL	
Range	Kelas
677 - 604	A
603 - 531	B
530 - 457	C
456 - 384	D
< 384	E





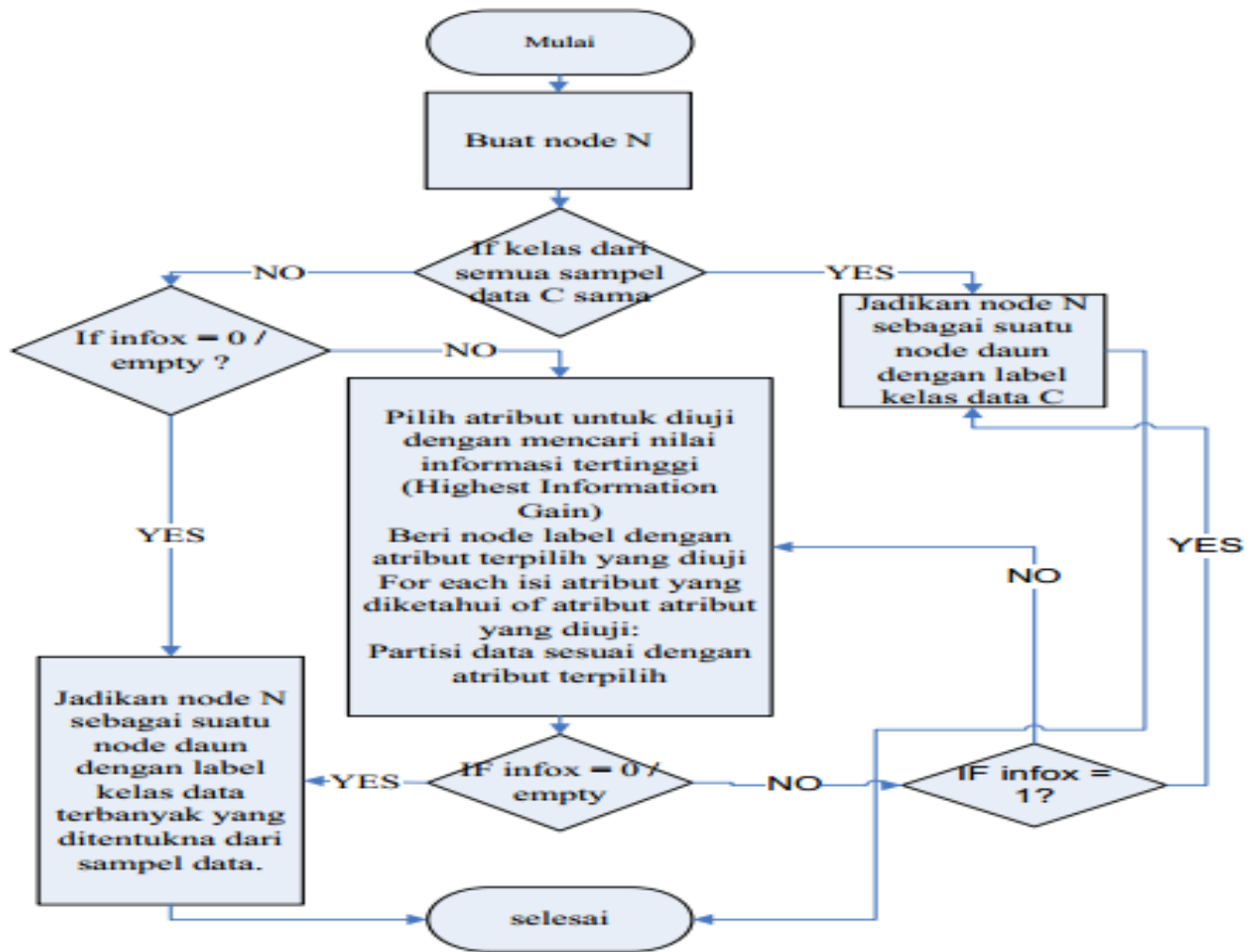
# Classification



- Proses untuk menyatakan suatu objek ke salah satu kategori yg sudah didefinisikan sebelumnya.



# ALGORITMA C.45







# Perhitungan Gain dan Entrophy

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Dengan :

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S<sub>i</sub>| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

dengan :

S : Himpunan Kasus

A : Fitur

n : Jumlah partisi S

p<sub>i</sub> : Proporsi dari S<sub>i</sub> terhadap S

# HTML, CSS, PHP dan MySql

- HyperText Markup Language (HTML) : sebuah bahasa markup untuk membuat halaman web.
- CSS (Cascading Style Sheets) : memungkinkan pemisahan antara konten dokumen.
- PHP adalah singkatan dari "PHP: Hypertext Preprocessor", yaitu bahasa pemrograman untuk pembuatan dan pengembangan sebuah situs Web, digunakan bersamaan dengan HTML
- MySQL adalah sebuah perangkat lunak sistem manajemen basis data SQL (database management system)







# Recall, Precision, dan Accuracy

		Nilai sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) <i>Corect result</i>	FP (False Positive) <i>Unexpected result</i>
	FALSE	FN (False Negative) <i>Missing result</i>	TN (True Negative) <i>Corect absence of result</i>

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$



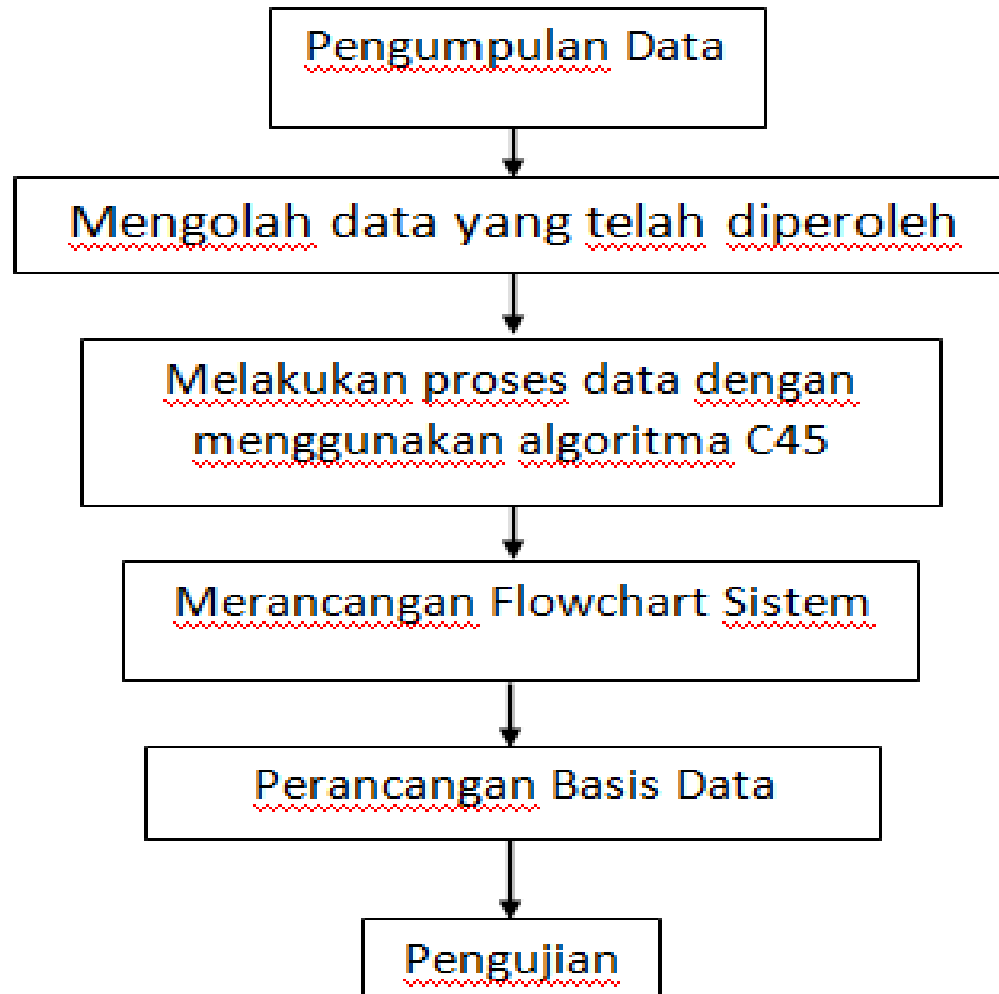
**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# METODOLOGI PENELITIAN



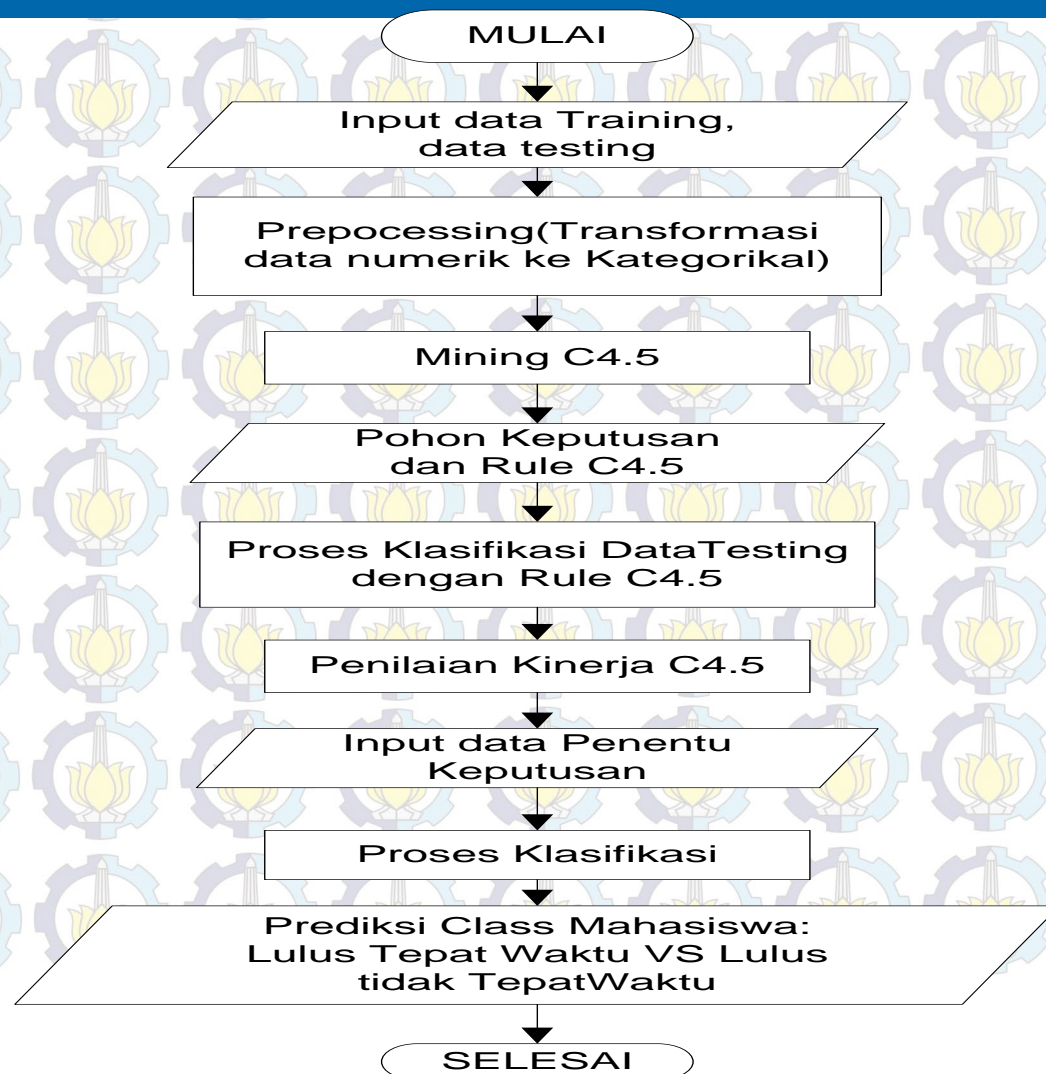


# Metodologi





# Flowchart System





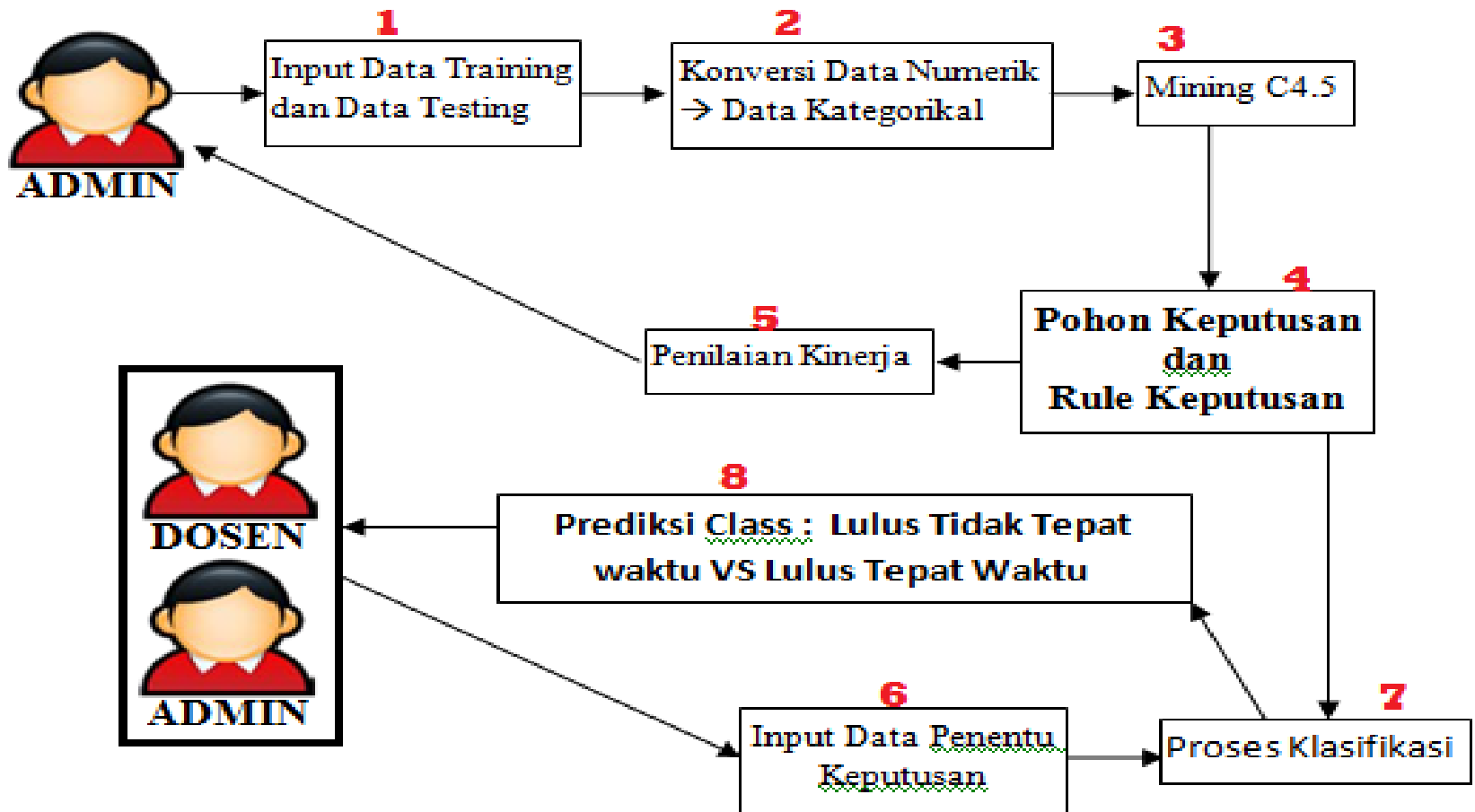


# Data Penelitian

- Penelitian ini menggunakan data mahasiswa lulusan MMT – ITS dari angkatan 2009-2012
- Variabel yang diteliti : Akreditasi S1, Program Studi, Status Program Studi S1, Bidang Keahlian, Biaya, IPK S1 ,GMAT , TOEFL, MBid, Wawancara, Skor Akhir, IPS Semester 1, IPS Semester 2, Jarak S1-S2.



# ALUR KERJA







**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# UJICoba SISTEM



# UJICoba SISTEM

- Ujicoba sistem dilakukan dalam 3 skenario:
  - Skenario I : Saat Mahasiswa baru diterima sebagai mahasiswa MMT ITS
  - Skenario II: Saat Mahasiswa telah menyelesaikan semester I
  - Skenario III : Saat Mahasiswa telah menyelesaikan semester II





# HASIL SKENARIO I

## Precision

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	33.33%	37.00%	60.00%	28.57%	25.00%	66.67%	41.76%
II	44.83%	40.00%	50.00%	55.56%	50.00%	40.00%	46.73%
III	33.33%	28.57%	66.67%	75.00%	50.00%	100.00%	58.93%
IV	42.86%	30.00%	40.00%	50.00%	26.67%	50.00%	39.92%
V	27.27%	26.67%	50.00%	40.00%	20.00%	100.00%	43.99%
AVG	36.32%	32.45%	53.33%	49.83%	34.33%	71.33%	46.27%

## Recall

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	31.58%	27.27%	37.50%	21.05%	18.18%	25.00%	26.76%
II	61.90%	46.15%	50.00%	23.81%	15.38%	33.33%	38.43%
III	12.00%	33.30%	25.00%	12.00%	12.50%	12.50%	17.88%
IV	50.00%	75.00%	40.00%	11.11%	50.00%	20.00%	41.02%
V	37.50%	30.77%	60.00%	12.50%	7.69%	40.00%	31.41%
AVG	38.60%	42.50%	42.50%	16.09%	20.75%	26.17%	31.10%

## Accuracy

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	52.83%	63.89%	61.11%	52.83%	58.33%	61.11%	58.35%
II	54.72%	55.56%	66.67%	62.26%	63.89%	61.11%	60.70%
III	47.17%	50.00%	61.11%	56.60%	66.67%	61.11%	57.11%
IV	60.38%	55.56%	66.67%	66.04%	58.33%	72.22%	63.20%
V	50.94%	44.44%	72.22%	67.92%	55.56%	83.30%	62.40%
AVG	53.21%	53.89%	65.56%	61.13%	60.56%	67.77%	60.35%



# HASIL SKENARIO II

## Precision

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	36.84%	30.00%	71.43%	33.33%	25.00%	66.67%	43.88%
II	46.15%	41.67%	50.00%	55.56%	50.00%	40.00%	47.23%
III	40.00%	36.36%	66.67%	75.00%	50.00%	100.00%	61.34%
IV	44.00%	33.33%	33.30%	50.00%	26.67%	50.00%	39.55%
V	27.27%	28.57%	50.00%	40.00%	20.00%	100.00%	44.31%
AVG	38.85%	33.99%	54.28%	50.78%	34.33%	71.33%	47.26%

## RECALL

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	36.84%	27.27%	62.50%	26.32%	18.18%	25.00%	32.69%
II	57.14%	38.46%	50.00%	23.81%	15.38%	33.33%	36.35%
III	16.00%	33.33%	25.00%	12.00%	16.67%	12.50%	19.25%
IV	61.11%	75.00%	40.00%	11.11%	56.00%	20.00%	43.87%
V	37.50%	30.77%	40.00%	12.50%	7.69%	40.00%	28.08%
AVG	41.72%	40.97%	43.50%	17.15%	22.78%	26.17%	32.05%

## ACCURACY

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	54.72%	58.33%	72.22%	54.72%	58.33%	61.11%	59.91%
II	56.60%	58.33%	66.67%	62.26%	63.89%	61.11%	61.48%
III	49.06%	58.33%	61.10%	56.60%	66.67%	61.11%	58.81%
IV	60.38%	61.11%	61.11%	66.04%	58.33%	72.22%	63.20%
V	50.94%	47.22%	72.22%	67.92%	55.56%	83.33%	62.87%
AVG	54.34%	56.66%	66.66%	61.51%	60.56%	67.78%	61.25%





**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# HASIL SKENARIO III

## Precision

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	40.91%	9.09%	57.14%	33.33%	25.00%	57.14%	37.10%
II	48.15%	41.67%	42.86%	55.56%	50.00%	40.00%	46.37%
III	41.67%	45.95%	75.00%	75.00%	50.00%	100.00%	64.60%
IV	42.86%	31.58%	42.86%	50.00%	26.67%	50.00%	40.66%
V	31.82%	44.00%	60.00%	40.00%	20.00%	100.00%	49.30%
<b>AVG</b>	<b>41.08%</b>	<b>34.46%</b>	<b>55.57%</b>	<b>50.78%</b>	<b>34.33%</b>	<b>69.43%</b>	<b>47.61%</b>

## recall

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	47.37%	9.09%	50.00%	26.32%	18.18%	50.00%	33.49%
II	61.90%	38.46%	50.00%	23.81%	15.38%	33.33%	37.15%
III	20.00%	41.67%	37.50%	12.00%	16.67%	12.50%	23.39%
IV	50.00%	75.00%	60.00%	11.11%	56.00%	20.00%	45.35%
V	43.75%	61.54%	60.00%	12.50%	7.69%	40.00%	37.58%
<b>AVG</b>	<b>44.60%</b>	<b>45.15%</b>	<b>51.50%</b>	<b>17.15%</b>	<b>22.78%</b>	<b>31.17%</b>	<b>35.39%</b>

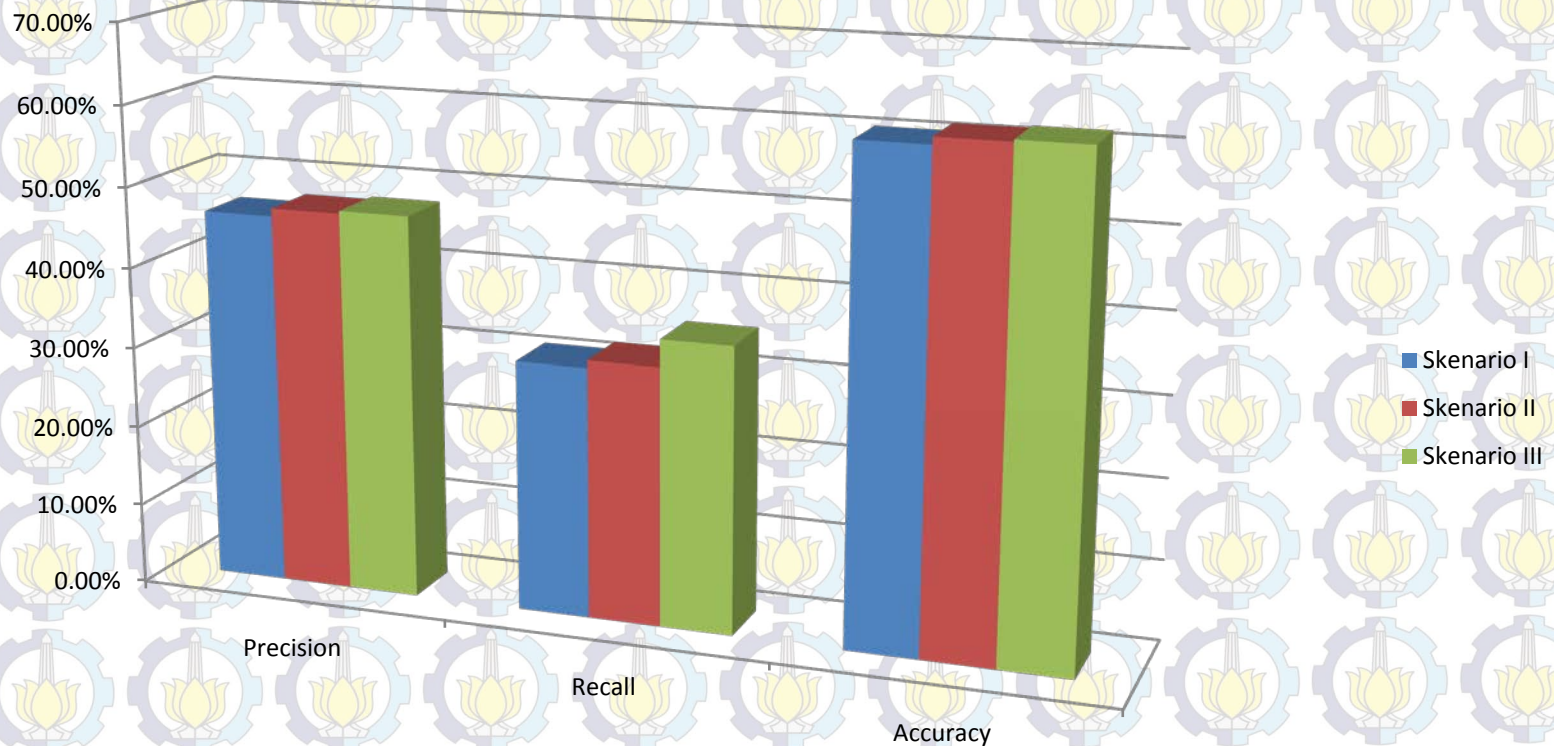
## Accuracy

UJI COBA	NON-PRUNING			PRE - PRUNING			AVG
	70:30	80:20	90:10	70:30	80:20	90:10	
I	56.60%	44.44%	61.11%	54.72%	58.33%	61.11%	56.05%
II	58.49%	58.33%	61.11%	62.26%	63.89%	61.11%	60.87%
III	49.06%	63.89%	66.67%	56.60%	66.67%	61.11%	60.67%
IV	60.38%	58.33%	66.67%	66.04%	58.33%	72.22%	63.66%
V	54.72%	58.33%	77.78%	67.92%	55.56%	83.33%	66.27%
<b>AVG</b>	<b>55.85%</b>	<b>56.66%</b>	<b>66.67%</b>	<b>61.51%</b>	<b>60.56%</b>	<b>67.78%</b>	<b>61.50%</b>



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# Perbandingan Skenario I, II dan III







# Pembahasan

- Proses Pre Pruning menyebabkan nilai accuracy meningkat, nilai precision meningkat dan nilai recall menurun.
- Bertambahnya data training menyebabkan nilai Accuracy, nilai Precision, dan nilai Recall meningkat.
- Nilai IPS 1 dan nilai IPS2 mempunyai peranan yang cukup signifikan terhadap identifikasi ketidak tepatan lulus mahasiswa dalam menyelesaikan studinya di MMT ITS.



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# Program Studi MTI (67 data)

SKN	Uji Coba	NON-PRUNING			PRE - PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
I	XII	40.00%	100.00%	57.14%	50.00%	100.00%	71.43%
	XIII	33.33%	100.00%	71.43%	25.00%	100.00%	57.14%
	XIV	33.33%	50.00%	57.14%	100.00%	50.00%	85.71%
	XV	33.33%	33.33%	42.86%	33.33%	33.33%	42.86%
	XVI	100.00%	25.00%	57.14%	50.00%	25.00%	42.86%
II	XII	33.33%	100.00%	42.86%	50.00%	100.00%	71.43%
	XIII	33.33%	100.00%	71.43%	25.00%	100.00%	57.14%
	XIV	33.33%	50.00%	57.14%	100.00%	50.00%	85.71%
	XV	33.33%	33.33%	42.86%	33.33%	33.33%	42.86%
	XVI	100.00%	25.00%	57.14%	50.00%	25.00%	42.86%
III	XII	40.00%	100.00%	57.14%	66.67%	100.00%	85.71%
	XIII	33.33%	100.00%	71.43%	25.00%	100.00%	57.14%
	XIV	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
	XV	33.33%	33.33%	42.86%	33.33%	33.33%	42.86%
	XVI	100.00%	25.00%	57.14%	50.00%	25.00%	42.86%
AVG		52.00%	61.67%	58.09%	52.78%	61.67%	60.95%



# Program Studi MI (73 data)



**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

SKN	Uji Coba	NON-PRUNING			PRE - PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
I	XII	33.33%	100.00%	71.43%	50.00%	100.00%	85.71%
	XIII	25.00%	50.00%	42.86%	50.00%	100.00%	85.71%
	XIV	33.33%	25.00%	28,57%	100.00%	25.00%	57.14%
	XV	50.00%	66.67%	57.14%	100.00%	33.33%	71.43%
	XVI	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
II	XII	33.33%	100.00%	71.43%	50.00%	100.00%	85.71%
	XIII	33.33%	50.00%	57.14%	100.00%	50.00%	85.71%
	XIV	50.00%	25.00%	42.86%	100.00%	25.00%	57.14%
	XV	50.00%	66.67%	57.14%	100.00%	33.33%	71.43%
	XVI	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
III	XII	25.00%	100.00%	57.14%	50.00%	100.00%	85.71%
	XIII	25.00%	50.00%	42.86%	100.00%	50.00%	85.71%
	XIV	50.00%	25.00%	42.86%	100.00%	25.00%	57.14%
	XV	50.00%	66.67%	57.14%	100.00%	33.33%	71.43%
	XVI	100.00%	50.00%	85.71%	100.00%	50.00%	85.71%
AVG		50.55%	58.33%	61.22%	86.67%	55.00%	77.14%

# Program Studi MP (33 data)

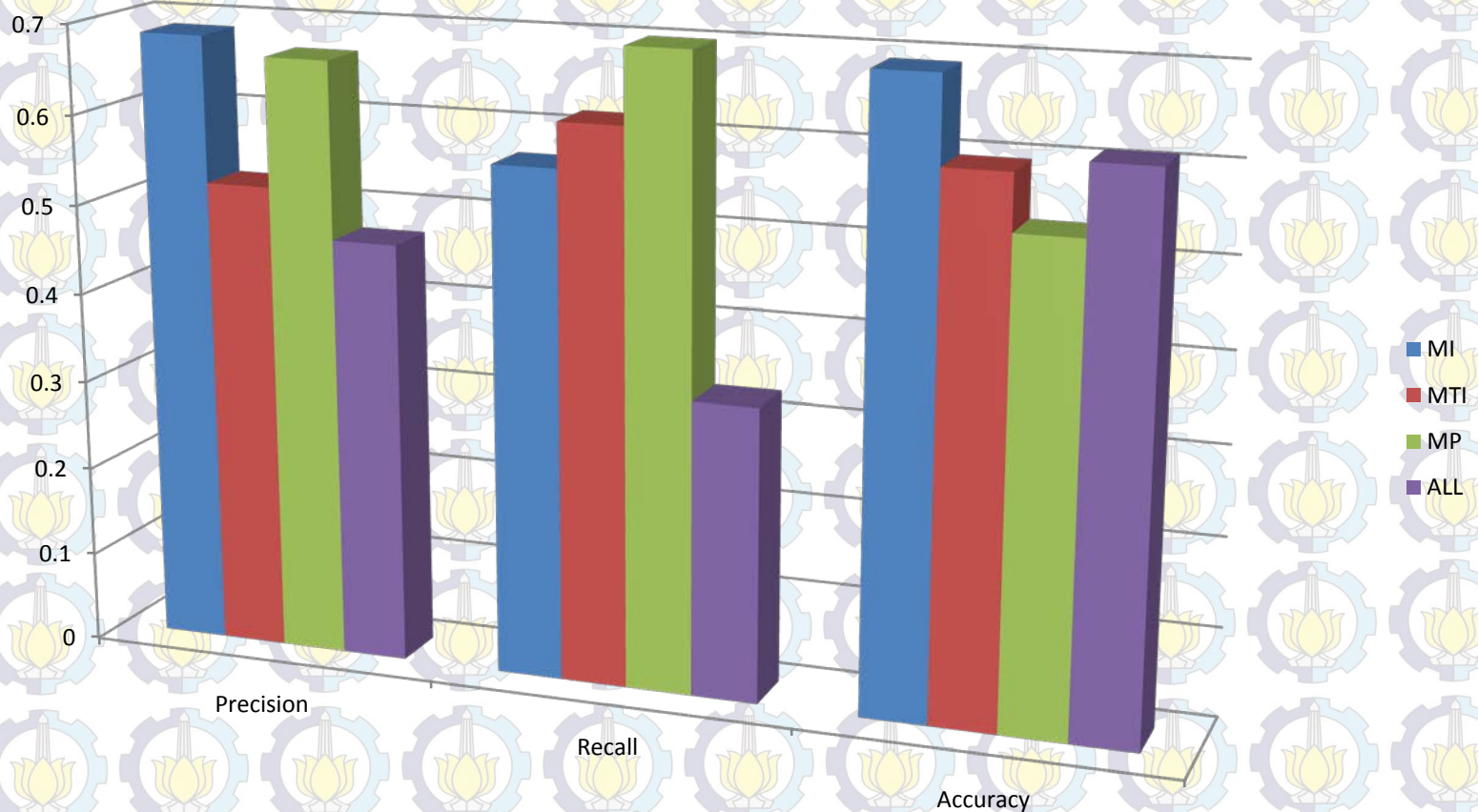
SKN	Uji Coba	NON-PRUNING			PRE - PRUNING		
		Precision	Recall	Acc.	Precision	Recall	Acc.
I	XII	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XIII	50.00%	50.00%	33.33%	50.00%	50.00%	33.33%
	XIV	33.33%	100.00%	33.33%	33.33%	100.00%	33.33%
	XV	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XVI	50.00%	100.00%	66.67%	50.00%	100.00%	66.67%
II	XII	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XIII	50.00%	50.00%	33.33%	50.00%	50.00%	33.33%
	XIV	33.33%	100.00%	33.33%	33.33%	100.00%	33.33%
	XV	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XVI	50.00%	100.00%	66.67%	50.00%	100.00%	66.67%
III	XII	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XIII	50.00%	50.00%	33.33%	50.00%	50.00%	33.33%
	XIV	33.33%	100.00%	33.33%	33.33%	100.00%	33.33%
	XV	100.00%	50.00%	66.67%	100.00%	50.00%	66.67%
	XVI	50.00%	100.00%	66.67%	50.00%	100.00%	66.67%
AVG		66.67%	70.00%	53.33%	66.67%	70.00%	53.33%





**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# Perbandingan Program Studi





# Pembahasan

- Nilai Precision meningkat bila dikelompokkan per program studi. Hal ini membuktikan bahwa mahasiswa pada program studi akan banyak memiliki kesamaan, mengelompok.
- Nilai Recall Meningkat bila dikelompokkan per program studi. Hal ini membuktikan bahwa ketika rule yang dipakai untuk uji coba khusus terhadap program studi itu maka jangkauan sistem dapat memprediksi lebih tepat.





# Ilustrasi Kemanfaatan Sistem

- Apabila dalam satu angkatan MTI terdapat 130 mahasiswa. Bila sistem memprediksikan ada 50 orang yang cenderung lulus tidak tepat waktu pada awal studinya maka:
- Acuan nilai precision: 48 %
- Maka yang tidak lulus tepat waktu sebenarnya :  $48\% = x/50$
- Sehingga  $x = 24$  mahasiswa yang lulus tidak tepat waktu.



# Ilustrasi Kemanfaatan Sistem

- Dari 50 siswa yang teridentifikasi kecenderungan lulus tidak tepat waktu setidaknya 24 orang siswa diantara dapat dipastikan lulus tidak tepat waktu.
- MMT ITS setidaknya bisa menyelamatkan  $24/130 * 100 \% = 18\%$  mahasiswanya.





**ITS**  
Institut  
Teknologi  
Sepuluh Nopember

# TERIMA KASIH